

Characterizing Entropy in Statistical Physics and in Quantum Information Theory

Bernhard Baumgartner¹

Institut für Theoretische Physik, Universität Wien
Boltzmanngasse 5, A-1090 Vienna, Austria

June 25, 2012

Abstract

A new axiomatic characterization with a minimum of conditions for entropy as a function on the set of states in quantum mechanics is presented. Traditionally unspoken assumptions are unveiled and replaced by proven consequences of the axioms. First the Boltzmann-Planck formula is derived. Building on this formula, using the Law of Large Numbers - a basic theorem of probability theory - the von Neumann formula is deduced. Axioms used in older theories on the foundations are now derived facts.

PACS numbers: 02.50.Cw, 03.67.Hk, 05.30
MSC code: 94A17

Keywords: Entropy, axiomatic, information, large numbers

¹Bernhard.Baumgartner@univie.ac.at

1 Introduction

The formula for the entropy of a density matrix,

$$S(\rho) = S_{vN}(\rho) = -k_B \text{Tr} \rho \ln(\rho), \quad (1)$$

introduced by von Neumann [vN29], has been successfully used in statistical mechanics dealing with equilibrium states of large systems. The enormous success of its use has the effect, that questions about a basic justification have been posed only lately. The same is to be said about Shannon's formula, put forward in [S48, SW49] for the entropy of classical information; but see [AFN74].

We see several reasons to “revisit” the basic axiomatics. One is to ask for counterparts to Lieb-Yngvason's axioms for thermodynamic entropy and the second law of thermodynamics, [LY98, LY99, LY00, LY03]. Then the development of quantum engineering, handling of nano-physics and emergence of quantum-information leads to a new interest in basic theories of entropy-like functionals, characterizing states according to their “mixedness”, giving some kind of distance to the pure states. This characterization needs not a priori to be in relation to thermodynamics.

Not only many aims are new, also the foundations both of quantum mechanics and of thermodynamics are now seen differently, in another light than in former times. Entanglement and decoherence may be considered as a special origin of the second law of thermodynamics, as is done f.e. in [PSW06], and explanations of concepts by referring to our “natural” classical intuition, as in [O75] which follows on [AFN74], seem questionable. These classical notions refer to considerations on building up a system from its parts. But the present paper is built on the basic concepts of probability as expressed in density matrices. The fundamental relation of probability to the physical world, irrespective of philosophical interpretations, is through the law of large numbers. Contrary to considering a state of a system as made up of different parts we consider unlimited repetitive appearance of one and the same state, drawing a characterization of the single state out of its appearance in large numbers. After the presentation of formal settings and of the axioms in the two following sections, a discussion of deeper thoughts on epistemological considerations and reasoning follows. Also the backgrounds in history and in more recent developments are included in this discourse.

A word on “axioms”: A set of axioms can either define a framework for a whole class of theories, (like Kolmogoroff's axioms for probability), or it can, in the other extreme, present the postulates building up one special theory (like Newton's axioms for mechanics or Peano's axioms for the natural numbers). While the Lieb-Yngvason axioms are rather of the first kind, the axioms presented here are a sequence of increasingly detailed postulates leading in several steps to the von Neumann entropy.

2 The axioms and their framework

2.1 Setup

Non-relativistic quantum mechanics is the framework. Each physical system \mathcal{A} , \mathcal{B} , ... is represented by a separable Hilbert space $\mathcal{H}_{\mathcal{A}}$, $\mathcal{H}_{\mathcal{B}}$, ... , states are represented by density matrices $\rho_{\mathcal{A}}$, $\rho_{\mathcal{B}}$, Superselection rules like conservation of particle numbers give a decomposition into superselection sectors $\mathcal{H}_{\mathcal{A}} = \bigoplus_{\ell} \mathcal{H}_{\mathcal{A},\ell}$, with restricting density matrices to be of block-diagonal form, $\rho_{\mathcal{A}} = \bigoplus_{\ell} \rho_{\mathcal{A},\ell}$, forbidding non-zero off-diagonal matrix elements connecting different sectors.

A compound system $\mathcal{A} \cup \mathcal{B}$ consisting of distinguishable subsystems \mathcal{A} , \mathcal{B} is represented by $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$. A superselection sector for system \mathcal{A} is defined as a space spanned by all the eigenvectors belonging to an eigenvalue q of an operator $Q_{\mathcal{A}}$, or to a set of eigenvalues q_{α} of several operators $Q_{\alpha,\mathcal{A}}$, like particle numbers and charges. In a compound system the particles, charges, etc. may be distributed over the subsystems. A sector there is defined as a subspace characterized by the eigenvalues of $Q_{\alpha} = Q_{\alpha,\mathcal{A}} \otimes \mathbb{1} \oplus \mathbb{1} \otimes Q_{\alpha,\mathcal{B}}$.

Considering compound systems consisting of several versions of the same system, we denote the n-fold tensor products as $\mathcal{H}_{\mathcal{A}}^{\otimes n}$. If each one of the subsystems is in the same state, and when there are no correlations between them, we denote the composed state as $\rho_{\mathcal{A}}^{\otimes n}$.

In the course of stating axioms and developing their implications we use firstly systems with Hilbert spaces where all superselection numbers are specified, like “a box with three atoms of gold and one atom of lead”. I name such a space an “undivided” Hilbert space. Considering transformations of states, either through natural or technically manipulated time evolution, or as hypothetical reversible mappings, each unitary transformation $\rho \mapsto \sigma = U \rho U^*$ on such an undivided Hilbert space is admissible. Acting in a Hilbert space with several sectors, to be consistent with the superselection rules the unitaries have to commute with all the Q_{α} . From standard quantum theories of particles we take the existence of an undivided Hilbert space with infinite dimension for granted.

We distinguish a special type of density matrix $\pi(N) := \frac{1}{N} P_N$, with P_N a projector onto an undivided (sub)space with dimension N . Such a $\pi(N)$ is the quantized version of distributions introduced by Laplace in the theory of chance and by Boltzmann in Statistical Mechanics; we call these $\pi(N)$ the “Quantum Laplace Boltzmann (QLB) states”.

When it comes to the details concerning the law of large numbers, the partial traces are needed to form the essential one-system density matrices ω defined on \mathcal{H} , out of a n-system density matrix Ω defined on $\mathcal{H}^{\otimes n}$.

$$\omega(k) := \text{Tr}_{\text{partial}, k \text{ excluded}}(\Omega) \quad (2)$$

Finally, to get the best possible kind of continuity (which will turn out to be only a lower semicontinuity), we refer to the topology of states induced by the trace

norm of density matrices.

Now the **entropy** S shall be a non-negative function on the set of states. (A discussion on this simplifying assumption is presented in Section 3.) This function S shall obey the following

2.2 Axioms

2.2.1 Equivalences

If a compound system is considered where one part of it is in a pure state like $\rho_B = |\psi_B\rangle\langle\psi_B|$, the entropy of the whole system is determined by the other part (or parts) only, as if part B were absent:

A) Decomposition: $S(\rho_A \otimes |\psi_B\rangle\langle\psi_B|) = S(\rho_A), \quad S(|\psi_A\rangle\langle\psi_A| \otimes \rho_B) = S(\rho_B).$

Each unitary transformation of states, which commutes with particle number operators and other superselection rules, conserves entropy:

B) Unitary invariance: $S(\rho_A) = S(\sigma_A) \quad \text{if} \quad \sigma_A = U \cdot \rho_A \cdot U^*, \quad [U, Q_\alpha] = 0 \quad \forall \alpha.$

These two axioms imply universality, they enable comparisons of all systems with undivided Hilbert spaces: One may choose one system with an infinite dimensional undivided Hilbert space \mathcal{H} (without an index) as a *reference* system, and a pure state $|\psi\rangle\langle\psi|$ with $\psi \in \mathcal{H}$. For any ρ_A on the undivided \mathcal{H}_A one has $S(\rho_A) = S(\rho_A \otimes |\psi\rangle\langle\psi|)$, and this compound state on $\mathcal{H}_A \otimes \mathcal{H}$, which is also undivided, may be unitarily transformed to $|\psi_A\rangle\langle\psi_A| \otimes \rho$ with any $\psi_A \in \mathcal{H}_A$. One gets $S(\rho_A) = S(\rho)$, now with ρ a density matrix on the referential Hilbert space \mathcal{H} , equivalent to ρ_A by a (partial) isometry. So S does not depend on the type of matter. One may compare the entropy of a system with three atoms of gold in one box with the entropy of two atoms of lead in another box. Entropy of such a state is a mathematical function of the spectral values (including multiplicity) of density operators only. It is also independent on the dimension of the Hilbert space.

2.2.2 QLB states and repeated appearances

Two QLB states $\pi(M)$ and $\pi(N)$ are equivalent in the sense of axioms A and B when $M = N$. Therefore their entropy is characterized by the dimension of their range only. It is “natural” to consider here strict monotonicity:

C) Monotonicity: $M > N \Rightarrow S(\pi(M)) > S(\pi(N)).$

A compound system, consisting of n equivalent systems in completely equivalent states, without any correlation, shall have n times the entropy of each of

its parts:

D) Extensivity, discrete scaling: $S(\rho^{\otimes n}) = n \cdot S(\rho)$

Now we have excluded any trivial S , and we can begin to justify the restriction of S to non-negative values: Note that $\pi(1)$ is a pure state, by axiom A it is equivalent to $\pi(1) \otimes \pi(1)$, so discrete scaling gives

$$S(\pi(1)) = S(\pi(1) \otimes \pi(1)) = 2 \cdot S(\pi(1)).$$

Attribution of ∞ is impossible, since other $S(\pi(N))$ have to be strictly larger, $-\infty$ has been excluded a priori, so $S(\pi(1)) = 0$ remains as the only allowed value.

A “calibration” is possible, setting $S(\pi(2)) = k_B \cdot \ln 2$, and we can prove Boltzmann’s formula (written down by Planck) for the QLB states as a strict consequence of these axioms:

1 THEOREM. *The Boltzmann-Entropy.* *Assuming axioms A, ... D, and choosing some universal constant k_B , the Boltzmann-Planck formula for the entropy of QLB states holds:*

$$S(\pi(N)) = k_B \cdot \ln N. \quad (3)$$

Proof. Observe that $\pi(N)^{\otimes n} = \pi(N^n)$, and consider numbers m, n , obeying $2^m \leq N^n < 2^{m+1}$. Then $m \cdot k_B \cdot \ln 2 \leq n \cdot S(\pi(N)) < (m+1) \cdot k_B \cdot \ln 2$. In the limit of large numbers, N fixed, one gets

$$S(\pi(N)) = \left(\lim_{n \rightarrow \infty} \frac{m}{n} \right) \cdot k_B \cdot \ln 2 = k_B \cdot \ln N.$$

□

2.2.3 Generality via Large Numbers

The properties of entropy shall reflect the law of large numbers. It states that repeated appearance of the same state, without correlations, becomes similar to a certain compound state with discrete uniform distribution. In subsection 3.4 we present details, justifying the following strict condition on the function S .

E) Rational combinatorics with large numbers:

If the density matrix ρ has finite rank and only rational numbers as eigenvalues, consider numbers n which are common multiples of their denominators and consider QLB states Ω with range in $\mathcal{H}^{\otimes n}$ which simulate the state $\rho^{\otimes n}$ in such a way, that each partial trace $\omega(k)$ of Ω , as defined in equ.(2), gives exactly ρ . The entropy of $\rho^{\otimes n}$ which, by axiom D, is n times the entropy of ρ , shall dominate the entropy of each of these Ω , but it shall not be larger then necessary for this dominance:

$$S(\rho) := \sup_{n, \Omega} \left\{ \frac{1}{n} S(\Omega) \mid \text{range}(\Omega) \subset \mathcal{H}^{\otimes n}, \Omega \text{ a QLB state}, \forall k \omega(k) = \rho \right\}$$

Those states for which entropy is already defined through axioms A ... E form a dense subset of the set of states for each system. On Hilbert spaces with finite dimension the remaining infinitesimal gaps can be filled in in such a way, that S becomes a continuous function. In case of \mathcal{H} being of infinite dimension the entropy S can at best only be extended to a lower semicontinuous function, including $+\infty$ in its range.

F) Semicontinuity: $S(\rho)$ is lower semicontinuous.

$$S(\rho) := \inf_{\sigma(n)} \left\{ \lim_{n \rightarrow \infty} S(\sigma(n)) \mid \sigma(n) \rightarrow \rho \text{ in trace norm, as } n \rightarrow \infty \right\}$$

2 THEOREM. von Neumann-Entropy. *The only function fulfilling all the axioms is von Neumann's entropy, as written in (1).*

Proof. a) $S_{vN}(\rho)$ is an upper bound for the entropy $S(\rho)$ of each ρ considered in E: The range of Ω is, by the conditions on the partial traces, a subset of $(\text{range}(\rho))^{\otimes n}$ which equals $\text{range}(\rho^{\otimes n})$. For simplicity we restrict the Hilbert space \mathcal{H} to $\text{range}(\rho)$. There we define the operator L_1 and on $\mathcal{H}^{\otimes n}$ its "second quantized form" L ,

$$L_1 := \ln \rho, \quad L := \sum_k L_{1,k}, \quad (4)$$

where $L_{1,k}$ acts in the k^{th} factor of $\mathcal{H}^{\otimes n}$. The repeated appearance of ρ in every subsystem, $\forall k \in \{1 \dots n\}$: $\omega(k) = \rho$, implies

$$\text{Tr}(\Omega \cdot L) = n \cdot \text{Tr}(\rho \cdot L_1) = \text{Tr}(\rho^{\otimes n} \cdot L). \quad (5)$$

By Klein's inequality and the *maximum entropy principle*, [W78, W90], this implies $S_{vN}(\rho^{\otimes n}) \geq S_{vN}(\Omega)$. We present a proof, adapted to the present situation: Klein's inequality leads to positivity of relative entropy (section I,B5 in [W78], or 3.1 in [R02], or [NC00]):

$$\text{Tr}(\Omega(\ln \Omega - \ln \rho^{\otimes n})) \geq 0.$$

Note that (4) implies $L = \ln \rho^{\otimes n}$, and therefore, using also (5):

$$\begin{aligned} S_{vN}(\Omega) &= -\text{Tr}(\Omega \ln \Omega) \leq \\ &\leq -\text{Tr}(\Omega \cdot \ln \rho^{\otimes n}) = -\text{Tr}(\Omega L) = -\text{Tr}(\rho^{\otimes n} L) = -\text{Tr}(\rho^{\otimes n} \ln \rho^{\otimes n}) = S_{vN}(\rho^{\otimes n}) \\ \text{Now } S(\Omega) &= S_{vN}(\Omega) \text{ and } S_{vN}(\rho^{\otimes n}) = n \cdot S_{vN}(\rho). \end{aligned}$$

b) For each ρ considered in E the supremum is reached in the limit $n \rightarrow \infty$, giving $S_{vN}(\rho)$: Consider the limit of those large numbers n which are multiples of a common denominator of all the eigenvalues r_j of ρ . Define Ω as the state proportional to the projector onto the linear span of those product states formed with eigenstates ϕ_j of L_1 , where each such ϕ_j appears m_j times, where $m_j = n \cdot r_j$.

Each eigenvector ϕ_j of ρ can be chosen as an element of a superselection sector, an eigenvector of all those operators Q_α whose spectral projectors define the sectors. With $Q_\alpha|\phi_j\rangle = q_{\alpha,j}|\phi_j\rangle$ each of the chosen product states has the same eigenvalues

$$Q|\Omega\rangle = \left(\bigoplus_k Q_{\alpha,k}\right) \bigotimes_k |\phi_{j(k)}\rangle = \sum_j m_j q_{\alpha,j}.$$

These product states are all in the same sector of $\mathcal{H}^{\otimes n}$, so the constructed density matrix Ω is one block, it is a QLB state for which the axioms A ... D are valid.

Combinatorics gives $\dim(\text{range}(\Omega)) = N$, with

$$N = \binom{n}{m_1! \cdot m_2! \dots m_\ell!}. \quad (6)$$

By Stirling's formula, the limit $n \rightarrow \infty$ gives $\lim_{n \rightarrow \infty} \frac{1}{n} \ln N = -\sum_j m_j \ln(m_j)$.

c) *Filling the infinitesimal gaps*: The ρ considered in axiom E form a dense subset in the set of all density matrices. On this subset the entropy takes on finite values. But if $\dim(\mathcal{H}) = \infty$, the values are unbounded, and not continuous: For $\rho = \sum_j r_j \cdot |\phi_j\rangle\langle\phi_j|$ consider the sequence $\rho(N) = \rho - \frac{1}{N}|\phi_1\rangle\langle\phi_1| + \frac{1}{N}\pi(N^2)$, with $N \geq 1/r_1$. The sequence of von Neumann entropies of $\rho(N)$ diverges. So von Neumann entropy, which is known to be lower semicontinuous, can at best be extended as a semicontinuous function, including $+\infty$ in its range. \square

3 Discussion

3.1 On the setup and on history

How did the founding fathers proceed, and is there a relation between their thoughts and the present point of view? Of the fathers of Statistical Physics, Clausius, Maxwell, Boltzmann (at the time of the precursor Daniel Bernoulli a concept of entropy was not yet in existence), Boltzmann was the first one who tried to give an expression for thermodynamic entropy by atomistic statistical terms. It is not easy to follow his thoughts in depth, see [K73, GRY08]. But he, and then Gibbs, Planck, Einstein and von Neumann, [B71, B72, B77, G02, P01, P06, E14, vN29], each one of them gave a justification of his theory by demonstrating its applicability in thermodynamics, mainly concerning an ideal gas. Even in von Neumann's book on the mathematical foundations of quantum

mechanics, [vN32], entropy as a function of density matrices is presented only in the subchapter V2 “Thermodynamische Betrachtungen”.

Now the usefulness of their atomistic formulas giving entropy in Statistical Physics is without any doubt. But, in the hindsight, one may question the uniqueness: Is it possible to find another formula, working as well as the existing one? That’s one of the questions where the analysis presented in this paper may help to find answers. Other questions may be asked in the emerging theory of Quantum Information. There is no inductive derivation of a quantity like entropy of quantum information from experiments and practice. But one can state “desiderata” how to characterize an appropriate measure for information carried by quantum states. I guess, they should be identical to those posed in Statistical Physics.

We deal with general states in non-relativistic Quantum Mechanics. Entropy has in some way to indicate the “mixedness” of a mixed state. As is well known, quantum mechanical mixing is in general different from the classical mixing of probability distributions, in that the decomposition into pure states is not unique. But there are still cases of classical mixing appearing in quantum mechanics. Consider for example a single point in a lattice gas of fermions. There are only two pure states; the lattice point is either empty or occupied. No coherent superposition is allowed. Quantumness appears only in a larger set of lattice points, where there exist pure states of a particle with a wave function which extends over several points. But, however large the system may be, there exists no coherent superposition of states with different numbers of particles. Such a superselection rule introduces a classicality into Quantum Physics. So, a priori, we have here to distinguish a fully quantum mechanical system with all mixed states out of the same superselection sector, from states where different sectors are involved in the mixing. It will come only through considerations of large numbers of parts, forming one large system, that a universal formula for entropy emerges, making no difference between the kind of mixing. The a priori difference has to be considered when postulating the invariance of entropy under unitary transformations, done in the following subsection 3.2.

In the vast universe of probability distributions and of mixed quantum states there appears a special kind of mixing: Laplace, in establishing basic methods for dealing with probability, considered sets of events with equal probabilities, because of a principle of insufficient reason, later named “the principle of indifference”. Boltzmann, considering distribution functions on phase space, performing a careful analysis of compatibility with Hamilton-Jacobi theory of classical mechanics, used *equal a priori probability* distributions on sets of given energy. Now, a QLB state $\pi(N)$ is the quantized version of such a distribution.

In analogy to the classical theories, one may interpret a QLB state as “each pure state characterized by a $\psi \in \text{range}(\pi(N))$ has the same probability”. We don’t want to build on special interpretations, we use the QLB states just as special states, discerned from others by their symmetry. This symmetry makes

handling them extremely simple. But that is again a case where we have to consider superselection rules. Mixed states with contributions from different sectors do not allow for as many decompositions into pure states as do mixed states which live in only one sector. It has less symmetry. So we start with these special QLB states with the highest possible symmetry.

Regarding in a short detour extreme philosophical considerations, whether the world should be modeled with features of infinity, or not, there appears the case that we should consider models of a world where only sectors with finite dimension exist. Now even if the elementary systems have only one-dimensional sectors, as in the example of a lattice gas with fermions, quantum mechanics comes into play. This happens through extensions in the compound systems. One just has to exclude that the world is completely trivial, being only one-dimensional. We could actually perform our considerations with a model of the world where there exist only sectors of finite dimension, but without an upper bound. But, for simplicity, we assume the existence of a sector with infinite dimension. This sector can serve as the reference Hilbert space.

In classical mechanics entropy can take on all values, between minus and plus infinity and including these values. The desiderata, as expressed in the axioms, do actually lead to strict positivity of the quantum mechanical entropy function, with the exception of pure states, see subsection 3.3. Again for simplicity we start already with demanding non-negativity of $S(\rho)$.

3.2 Equivalences, the way from physics to mathematics

Decompositions of systems are, at least since Galileo Galilei, at the heart of modern physics. At this point we pose the minimal postulate, that parts of a compound system, which are in pure states, have no effect on the entropy, and may as well be absent.

Thinking of applications in thermodynamics, such a part can represent, in some simplification, a weight which can be heaved up or lowered in a gravitational field.

Comparability of states is an essential feature, it is the central point in the analysis of the second law and entropy in [LY99]. While Lieb and Yngvason postulate first the existence of an order relation between states and regard “adiabatic equivalence” as derived, ($\rho \sim \sigma$ if both $\rho \prec \sigma$ and $\sigma \prec \rho$), the procedure in the present paper starts with equivalences, axioms A and B. In this paper there is no explicit definition of a mathematical equivalence relation; talking now about equivalence of states, we refer to the applicability of these two axioms (which could be used for a precise definition).

Comparison, inequalities for the entropy function (axiom C), is then formulated for special equivalence classes. This relation of equivalence is the point, where unitary evolutions and transformations of states come into play. Already prior to Heisenberg’s and Schrödinger’s creation of Quantum Mechanics it has

been stated by Einstein and Szilard, that the time evolution of mixed states should be considered with the assumption that pure states evolve into pure states. Einstein, in [E14], calls this “Ehrenfests Adiabatenhypothese”; von Neumann in [vN29] cites this and also page 777 in [S24], where Szilard refers to his doctoral thesis, as precursors to his formulation of entropy.

At the time of creating quantum physics only time evolution as is given by nature had to be considered. Now we have the possibility of steering time evolution on an atomistic scale, [CTGO11], without observing contradictions to the Second Law of Thermodynamics. So we postulate that each unitary transformation which is consistent with superselection rules does not change entropy.

If we think of an eager freshman in physics with little a priori knowledge, we guess that this freshman would intend to think that states describing spins of electrons should carry completely different amounts of entropy than mathematically equivalent states describing spins of nuclei. But such a difference does not appear. Entropy is a mathematical property, independent of the kind of matter which is described by a density matrix. We regard this fact as remarkable and worth while to be derived from axioms.

Axiom B paves the way for comparison by establishing equivalences of states in one sector, as, f.e. states of three atoms of gold in one box. Combining now the equivalences of unitarily transformed states (axiom B) with forming and reducing compound systems by adding and discarding pure states of another system (axiom A) lays threads of equivalent states through all sectors. A system with three atoms of gold has states which are equivalent to states of a system with five atoms of lead, and equivalent to some states of a system with two atoms of gold in a box.

3.3 QLB states form the backbone

QLB states are completely characterized by one discrete parameter, so they show a truly “natural” ordering, allowing for a “natural” inequality between their entropies. This inequality can also be supported by considering certain irreversible evolutions in physics, namely “mixing”. If $M < N$, the QLB state $\pi(N)$ can be constructed by mixing several unitarily transformed $\pi(M)$. The irreversibility has to be represented in an increase of entropy. These ideas are the basis for Uhlmann’s order relation for density matrices, and could be used for a slightly different set of axioms, see the Appendix, section 4.

Axiom D, on strict rules for the entropy function in case of repeated uncorrelated appearance of parts, is the first step to the law of large numbers. It is here, that one could exclude negativity: $S(\pi(N^2)) = 2 \cdot S(\pi(N))$ and monotonicity imply $S(\pi(N)) > 0$. Only the case $S(\pi(0)) = -\infty$ would need to be excluded a priori or with an extra axiom.

Note that at this point the Boltzmann-Plank formula is derived for $\pi(N)$ states only, with the assumption that their range is in one sector. The proba-

bilistic distribution over different sectors has to be analyzed by using the law of large numbers.

Choosing the extensivity may be considered as convenient, but not absolutely necessary. For example, demanding $S(\rho^{\otimes n}) = (S(\rho))^n$ instead, would lead to a different, but closely related formula.

3.4 Flesh on the backbone by the law of large numbers

The procedure of deriving the $\rho \log \rho$ -formula from equal a priori distributions is not new. There are only differences in the settings. Boltzmann considers atoms in a gas, Gibbs shows the equivalence of canonical and microcanonical ensembles. Each of the founding fathers used expression (6) and Stirling's formula; and also did Schrödinger in [S46].

Here we consider states from the point of view of probability theory. Directly justifying the interpretations of entropy as a measure of uncertainty or of information is difficult. (This has been noted already in [BR81], chapter 6.2.3.) Nevertheless, entropy arises naturally in the formalism of the law of large numbers. This law is rarely addressed explicitly in the context of defining entropy, but it appears implicitly, when “average information” is mentioned (f.e. in [NC00]). It is basic to application of probability theory. (Jacob Bernoulli, who established the Law of Large Numbers, see f.e. [B01], regarded the derivation of this theorem as a greater achievement than if he had shown how to square a circle, since a proof the latter would have been of little use.)

This law says, roughly, that in a typical series of N measurements the number $n_i(N)$ of appearances of event i lies most probably in an interval

$$[N - c\sqrt{N}, N + c\sqrt{N}] \cdot \rho_i. \quad (7)$$

with ρ_i the probability for appearance of event i , and c some constant. Our axiom E seems, at a first glance, to deviate in two ways from this law. It accounts for a number of possible series which seems to be in one sense too small, in another sense too large. The number seems to be too small, since we restrict the $n_i(N)$ to one definite value, instead of letting it vary in the interval (7). But, taking the logarithm, this difference becomes negligible. Making a step from N to $2N$, all those series covered by the interval (7) appear in a series of doubled length, where $n_i(2N) = 2N\rho_i$ exactly. In the Boltzmann-Plank formula the logarithm is taken, and $\ln 2$ is negligible for large N .

On the other hand, the dimension of QLB-states appears as being too large, as quantum correlations between the repeated appearances are allowed. As an example consider a spin up - spin down probability distribution $2/3, 1/3$ and a series of $N = 3$ tests. In the 8-dimensional Hilbert space, a 3-dimensional subspace is accounted for by combinatorics, as in formula (6). It is spanned by the vector $\uparrow\uparrow\downarrow$ and its permuted analogues. The QLB state, according to the

conditions of axiom E, allows for one more state vector, namely $(2/3)^{1/2} \uparrow\uparrow\uparrow + e^{i\alpha}(1/3)^{1/2} \downarrow\downarrow\downarrow$, where α is an arbitrary phase. Since already the restriction of the $n_i(N)$ to strict values introduces a kind of classical correlation between different appearances, I see no reason to exclude a quantum correlation. But these extra states can not lead to an excess over the von Neumann entropy, as is shown in part *a* in the proof of Theorem 2.

4 Appendix: Uhlmann's order relation

States allow for a partial ordering related to mixing. It is Uhlmann's order relation, [U71, U72, U73, W74]. Note, that this order relation is not the ordering defined in [LY98]. In matrix analysis [B97] Uhlmann's order relation is known as majorization:

$\rho \succ \sigma$ means: ρ is more mixed than σ ; ρ is majorized by σ . The definition is

$$\rho \succ \sigma \quad \text{iff} \quad \forall N \quad \sum_{n=1}^N r_n \leq \sum_{n=1}^N s_n, \quad (8)$$

where the r_n and the s_n are the eigenvalues of ρ and of σ in decreasing order, $r_1 \geq r_2 \geq r_3 \dots$, and at least one of the inequalities is strict. (Note that $\pi(M)$ is more mixed than $\pi(N)$ if $M > N$.) With this ordering one could use different axioms:

C') Monotonicity: $\rho \succ \sigma \Rightarrow S(\rho) \geq S(\sigma)$, strict for σ with finite rank.

In matrix analysis the "monotonicity" is known as "Schur-concavity". Also this way we have excluded any trivial S , as above we conclude $S(\pi(1)) = 0$. All other states are more mixed and must have larger entropy. For states not of QLB form, there are still several possibilities, the first four axioms do f.e. not yet exclude Renyi-entropies. So it is again the law of large numbers which leads strictly to von Neumann's formula. The axiom F of semicontinuity can then be replaced by

F') : $S(\rho) := \sup_{\sigma} \{S(\sigma) \mid \rho \succ \sigma\}$.

References

- [AFN74] J. Aczél, B. Forte and C. T. Ng: *Why the Shannon and Hartley Entropies are 'Natural'*. Adv. Appl. Prob. **6**, 131 – 146, (1974)
- [B01] Daniel Brönnimann: *Die Entwicklung des Wahrscheinlichkeitsbegriffs von 1654 bis 1718*. [ftp://stat.ethz.ch/Masters.../Daniel Brönnimann-Wahrscheinlichkeit.pdf](ftp://stat.ethz.ch/Masters.../Daniel%20Brönnimann-Wahrscheinlichkeit.pdf)

- [B71] Ludwig Boltzmann: *Analytischer Beweis des zweiten Hauptsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht der lebendigen Kraft*. Sitzb. der Wiener Akad. **LXIII**, 712 – 732, (1871)
- [B72] Ludwig Boltzmann: *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*. Sitzb. der Wiener Akad. **LXVI**, 275 – 370, (1872)
- [B77] Ludwig Boltzmann: *Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht*. Sitzb. der Wiener Akad. **LXXVI**, 373 – 435, (1877)
- [B97] Rajendra Bhatia: *Matrix Analysis*. Springer, (1997)
- [BR81] Ola Bratteli, Derek W. Robinson: *Operator Algebras and Quantum Statistical Mechanics II*. Springer-Verlag, New York Heidelberg Berlin, (1981)
- [CT73] E.G.D. Cohen and Walter Thirring: *The Boltzmann Equation*. Springer-Verlag Wien-New York (1973)
- [CTGO11] Claude Cohen-Tannoudji, David Guéry-Odelin: *Advances in Atomic Physics*. World Scientific, (2011)
- [E14] Albert Einstein: *Beiträge zur Quantentheorie*. Verh. d. Deutsch. Phys. Ges. **12**, 820 – 828, (1914)
- [G02] J. Willard Gibbs: *Elementary Principles in Statistical Mechanics*. Yale University Press, (1902), republished by Dover Publ., Inc. N.Y., (1960)
- [GRY08] Giovanni Gallavotti, Wolfgang L. Reiter and Jakob Yngvason, Eds: *Boltzmann's Legacy*. European Mathematical Society Publ. House (2008)
- [K73] M. J. Klein: *The Development of Boltzmann's Statistical Ideas*. in [CT73]
- [LY98] Elliott H. Lieb and Jakob Yngvason: *A Guide to Entropy and the Second Law of Thermodynamics*. Notices of the Amer. Math. Soc. **45**, 571 – 581 (1999) Erratum **314**, 699 (1999)
- [LY99] Elliott H. Lieb and Jakob Yngvason: *The Physics and Mathematics of the Second Law of Thermodynamics*. Phys. Reports **310**, 1 – 96 (1999) Erratum **314**, 699 (1999)
- [LY00] Elliott H. Lieb and Jakob Yngvason: *A Fresh Look at Entropy and the Second Law of Thermodynamics*. Phys. Today **310**, 32 – 37 (2000)

- [LY03] Elliott H. Lieb and Jakob Yngvason: *The Mathematical Structure of the Second Law of Thermodynamics*. arXiv:math-ph/0204007
- [NC00] Michael A. Nielsen and Isaac L. Chuang: *Quantum Computation and Quantum Information*. Cambridge Inversity Press, Cambridge UK 2000
- [O75] W. Ochs: *A New Axiomatic Characterization of the von Neumann Entropy*. Rep. Math. Phys. **3**, 109 – 120, (1975)
- [P01] Max Planck: *Über das Gesetz der Energieverteilung im Normalspektrum*. Ann. d. Phys. **4**, 4, 553 – 563, (1901)
- [P06] Max Planck: *Vorlesungen über die Theorie der Wärmestrahlung*. Barth, Leipzig (1906)
- [PSW06] Sandu Popescu, Anthony J. Short and Andreas Winter: *Entanglement and the foundations of statistical mechanics*. Nature Phys. **2**, 754 – 758, (2006)
- [R02] Mary Beth Ruskai: *Inequalities for Quantum Entropy: A Review with Conditions for Equality*. J. Math. Phys. **43**, 4358 – 4375 (2002)
- [S24] Leo Szilard: *Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen*. Zeitschrift für Physik. **XXXII** (10), 753 – 788 (1925)
- [S46] Erwin Schrödinger: *Statistical Thermodynamics*. Cambridge University Press, London (1946)
- [S48] C. F. Shannon: *A Mathematical Theory of Communication*. The Bell System Technical Journal **27**, 379 – 423, 623 – 656, July, October (1948)
- [SW49] C. F. Shannon and W. Weaver: *The Mathematical Theory of Communication*. Univ. of Illinois Press, Chicago (1949)
- [U71] Armin Uhlmann: *Sätze über Dichtematrizen*. Wiss. Z. Karl-Marx-Univ. **20**, 633 – 637, (1971)
- [U72] Armin Uhlmann: *Endlich-dimensionale Dichtematrizen I*. Wiss. Z. Karl-Marx-Univ. **21**, 421 – 452, (1972)
- [U73] Armin Uhlmann: *Endlich-dimensionale Dichtematrizen II*. Wiss. Z. Karl-Marx-Univ. **22**, 139 – 177, (1973)
- [vN29] Johann von Neumann: *Thermodynamik quantenmechanischer Gesamtheiten*. Gött. Nachr. **1**, 273 – 291 (1929)

- [vN32] Johann von Neumann: *Mathematische Grundlagen der Quantenmechanik*. Springer-Verlag Berlin-Heidelberg, (1932)
- [W74] Alfred Wehrl: *How chaotic is a state of a quantum system?* Rep. Math. Phys. **6**, 15 – 28, (1974)
- [W78] Alfred Wehrl: *General properties of entropy*. Rev. Mod. Phys. **50**, 221 – 260, (1978)
- [W90] Alfred Wehrl: *Information-Theoretical Aspects of Quantum-Mechanical Entropy*. Univ. Vienna preprint UWThPh-1990-20, unpublished, (1990)